IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED DATA MINING METHODS FOR SOCIAL COMPUTING

Received July 20, 2019, accepted August 17, 2019, date of publication September 6, 2019, date of current version September 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939827

Channel Access and Power Control for Energy-Efficient Delay-Aware Heterogeneous Cellular Networks for Smart Grid Communications Using Deep Reinforcement Learning

FAUZUN ABDULLAH ASUHAIMI, (Student Member, IEEE), SHENGRONG BU, (Member, IEEE), PAULO VALENTE KLAINE, (Student Member, IEEE), AND MUHAMMAD ALI IMRAN, (Senior Member, IEEE)

Department of Electrical Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.

Corresponding author: Fauzun Abdullah Asuhaimi (f.abdullah-asuhaimi.1@research.gla.uk)

This work was supported in part by the DARE Project under Grant EP/P028764/1 under the Engineering and Physical Sciences Research Council's (EPSRC's) Global Challenges Research Fund (GCRF) allocation.

ABSTRACT Cellular technology with long-term evolution (LTE)-based standards is a preferable choice for smart grid neighborhood area networks due to its high availability and scalability. However, the integration of cellular networks and smart grid communications puts forth a significant challenge due to the simultaneous transmission of real-time smart grid data which could cause radio access network (RAN) congestions. Heterogeneous cellular networks (HetNets) have been proposed to improve the performance of LTE because HetNets can alleviate RAN congestions by off-loading access attempts from a macrocell to small cells. In this paper, we study energy efficiency and delay problems in HetNets for transmitting smart grid data with different delay requirements. We propose a distributed channel access and power control scheme, and develop a learning-based approach for the phasor measurement units (PMUs) to transmit data successfully by considering interference and signal-to-interference-plus-noise ratio (SINR) constraints. In particular, we exploit a deep reinforcement learning(DRL)-based method to train the PMUs to learn an optimal policy that maximizes the earned reward of successful transmissions without having knowledge on the system dynamics. Results show that the DRL approach obtains good performance without knowing the system dynamic beforehand and outperforms the Gittin index policy in different normal ratios, minimum SINR requirements and number of users in the cell.

INDEX TERMS Energy efficiency, end-to-end delay, device-to-device communications, cellular networks, smart grids.

I. INTRODUCTION

Smart grids have attracted a lot of attention due to their potential to significantly improve the efficiency and reliability of power grids [1]. The smart grids utilize bidirectional communications between various smart grid domains to coordinate energy generation, transmission and distribution, and the smart grid communications are an essential part of an efficient grid control [2]. In smart grids, the distribution levels are prone to faults caused by different situations, such as equipment errors and adverse weather [3], which might lead to service interruptions and power loss. Performance of the communication at distribution level is critical to ensure the stability of grids. Neighborhood area networks (NANs) hold communications at the distribution level, which involves transmitting meter and status data to the control center for various applications, such as demand-side management, distribution automation and outage management.

In smart grid, higher penetration of distributed energy resources (DERs) based on renewable energy such as solar and wind power planted at distribution level is expected in future associated with the rising of energy demand from user side [4]. The DERs are very dependent to local weather conditions and highly intermittent, which require additional

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Hajizadeh.

monitoring [5], therefore, in order to make monitoring and controlling possible at DERs at the distribution level, phasor measurement units (PMUs) are deployed. PMUs play a critical role to transmit real-time dynamic data on power flows to the power system control center [6]. PMU measurements are gained by first sampling the voltage and current waveforms through the global positioning system, then each sample is time-stamped for phase and amplitude variations assessment before it is sent to the local phasor data concentrator (PDC). Moreover, all PMUs in a microgrid are synchronized, i.e., measurements are transmitted to the PDC at the same time. Of the existing wireless technology, cellular technology with LTE-based standards is a good choice for NANs due to its high availability and flexibility [7]. However, the large volumes of simultaneous transmission of smart grid data from PMUs and other devices in NANs could cause severe RAN congestions, leading to excessive delay in conventional cellular networks [7], therefore HetNets are proposed as critical techniques to reduce RAN congestions. In HetNets, low-power base stations are exploited in a macrocell, which are located close to the edges of macrocell to improve the data rate of users. HetNets have the ability to alleviate RAN congestions by off-loading access attempts from macrocells to small cells [8].

Energy efficiency is one of the critical parameters in HetNets. When abnormal events occur, for instance natural disasters such as floods, earthquakes or tsunamis, the PMUs will be isolated from the grid. In this situation, the PMU is powered by local energy sources, such as small wind turbine, photovoltaic panels and local energy storage equipment [9], which have limited power supply. Therefore, energy efficiency is critical in this kind of situation to ensure that status of DERs can be transmitted to the control center successfully. However, increasing the energy efficiency might compromise delay, an important performance parameter that reflects the actual user experience in the network. Delay is also important for PMUs because if the PMU data exceed the delay requirements, information loss may occur, which might lead to power loss and blackouts may occur in severe cases [10]. Therefore, it is critical to consider both parameters in HetNets. Channel access and power control are two critical schemes in HetNets especially when energy efficiency and delay are considered. Channel access scheme can be exploited to satisfy the stringent delay requirements by allowing devices to properly select a communication channel that satisfies the quality-of-service (QoS) of their data. On the other hand, the power control scheme is one of the energy efficiency maximization schemes which permits transmission power regulation of devices with respect to some constraints. The combination of both schemes could result in better performance in HetNets.

Many studies have addressed the energy efficiency and delay problem in HetNets for cellular communications with different schemes. For example, Mohammad *et al*. proposed joint sub-carrier and power allocation scheme in energy harvesting-enabled-power domain non-orthogonal multiple access (PD-NOMA)-based HetNets and exploited optimal approach based on the monotonic optimization to solve the problem [11]. Karim *et al*. investigated cloud radio access network and ray tracing-based resource allocation problem in heterogeneous traffic LTE networks and adopted heuristic algorithms to cater the problem [12]. Lun *et al*. proposed joint user association, clustering, and on/off strategies in dense heterogeneous networks and exploited the semidefinite programming and effective approximation approach to obtain maximum energy efficiency with satisfied QoS [13]. Cong *et al*. exploited the on-line learning approach to solve the mobility management problem in highly dynamic ultra-dense HetNets [14].

In addition to works related to energy efficiency and delay in Hetnets, the increase in number of devices due to the integration of smart grid communications and cellular technology demands for self-organized communications in heterogeneous and massive system [15], and deep learning is an emerging tool which can be exploited. Deep learning can be defined as a class of machine learning algorithm in the form of a neural network that extracts features from data and make predictive guesses about new data using a cascade of layers of processing units. Deep reinforcement learning (DRL) combines reinforcement learning and deep learning, by exploiting deep neural networks method to develop an artificial agent that is able to learn optimal policies directly from high-dimensional sensory inputs using end-to-end reinforcement learning (RL) [16]. DRL is promising for wireless communication agents because DRL enables them to learn the system dynamics and obtain the optimal policy in random and dynamic environments without knowledge of the system [17]–[19]. Moreover, DRL has the ability to deal with high-dimensional and large system states such as in HetNets [20], [21]. Based on these reasons, DRL approach is exploited to train PMUs to access channel and regulate its power in order to achieve maximum energy efficiency and satisfy delay constraints in distributed manner by extracting inputs from the environment.

Although an extensive study has been conducted on energy efficient and delay in HetNets, all these works utilized conventional analytical optimization techniques and none of them tries to explore more intelligent algorithms involving deep learning. Moreover, none of them considered a joint channel access and power control scheme in HetNets to achieve the objectives by exploiting DRL. In this paper, we propose HetNets as a solution to reduce RAN congestion when devices in LTE attempt access simultaneously. In order to maximize energy efficiency and meet delay constraints of the PMUs in HetNets, we exploit an intelligent channel access and power control scheme by taking into account the differentiated delay requirements of the PMUs using a DRL approach. By adopting this approach, the PMUs can adapt with the varying wireless channel conditions [22] even without knowing the system dynamic beforehand, through interactions with the environment. Furthermore, historical data can be used to train the proposed algorithm, leading it

towards better decisions in the future and optimizing energy by considering the differentiated delay requirements due to different states of the DERs.

The contributions of this paper can be summarized as follows.

- We study energy efficiency and delay problems in HetNets by considering PMUs' data with different delay requirements.
- We propose channel access and power control schemes to achieve the objective for devices with high generation data in slow fading channel environment.
- We propose a DRL approach algorithm for distributed intelligent channel access and power control scheme in HetNets and analyze the distributed decision made by PMUs in a variety of different conditions.

The rest of this paper is organized as follows. Information on the DERs' state and delay requirements of PMUs data is provided in Section II. The system model is described in Section III. The DRL approach for intelligent channel access and power control scheme is explained in Section IV. Simulation results are presented and discussed in Section V and Section VI concludes the paper.

II. DERS' STATE AND DELAY REQUIREMENTS OF PMUS DATA

Most of the energy management system applications assume that the system is in a pseudo-steady state where alternating-current circuit analysis can be carried out using the PMUs. The PMUs, usually placed at the 24.9kV distribution lines in power networks, measure voltage and current phasors of DERs, and then directly compute real power and volt-ampere reactive (VAR) flows at precise moments [23], which is crucial for grid protection and monitoring. In general, the DERs can be operated in one of three states: normal state, abnormal state, and restorative state [24].

The DER is in a normal state when some component emergency ratings and the voltage can be maintained at a safe minimum, at the same time ensuring that the service to the control center can be maintained. When some of these components cannot be retained, the DER needs control commands from the control to move back to normal states. Assume that the DER moves from normal state to other states with probability ρ. Let $g_{i,t}$ denote the state of DER observed by PMU i at iteration t: normal (0), restorative (1) and abnormal (2). Depending on the $g_{i,t}$, the data from PMUs are used for different applications with different delay requirements. In normal states, the data measured by PMUs are used for controlling and monitoring applications with the delay requirement of 20 ms [25]. When abnormal events occur or the DER is in a restorative state, the data of PMUs are used for protection, in which delay delivery requirement is reduced to 8 ms [26]. If data exceed the delay requirements, information loss may occur, which might lead to power loss and blackouts may occur in severe cases.

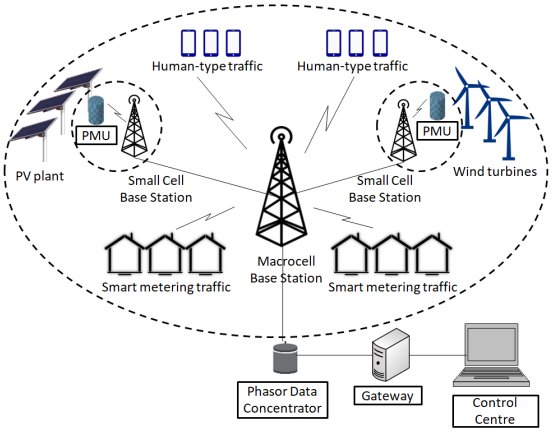


FIGURE 1. The architecture of heterogeneous cellular networks for the smart grid.

III. SYSTEM MODEL

A HetNet shown in Fig. 1 is considered for a smart grid NAN, where there is one macrocell base station (MBS) and E small cell base stations (SCBSs) underlaid on the macrocell and located close to the edge of the macrocell. These SCBSs are connected to the MBS through a wired network. The SCBSs offer traffic off-loading to improve the service rates. The communication devices in the macrocell include the PMUs, the smart meters and the mobile devices. All devices could attempt to access the network simultaneously. The PMUs, deployed close to the DERs, are responsible for collecting measurements related to the status of the DERs. The macrocell users (MUEs) are served by the MBS while the PMUs and the small cell users (SUEs) are served by the SCBSs for a higher service rate. The PMUs transmit the generated data to the PDC through the SCBSs and the MBS. After that, the PDC forwards the data to the control center to make decisions through the gateway in the core network.

The network is operated in a time-slot manner, where in each time slot, U sub-channels with the same bandwidth are licensed to the MBS. The MBS serves U MUEs by allocating one sub-channel to each MUE. At the same time, these U sub-channels are shared with the M PMUs and the J SUEs ($M < J \leq U$). The PMUs and the SUEs access the sub-channels intelligently in a distributed manner, and the MBS is aware of the spectrum accessed by these users. The set of the MUEs, the SCBSs, the PMUs, the SUEs and the sub-channels are denoted as $\mathcal{U} = \{1, \cdots, U\}$, $\mathcal{E} = \{1, \cdots, E\}$, $\mathcal{M} = \{1, \cdots, M\}$, $\mathcal{J} = \{1, \cdots, J\}$ and $\mathcal{N} = \{1, \cdots, N\}$ respectively.

Without any knowledge about the MUEs and the SUEs in the cell, the PMUs access the sub-channels and regulate their transmission power to maximize their own reward. Regardless of this greedy behavior, it is important for the PMUs to adapt to the environmental changes as energy efficiency is highly dependent on environmental factors such, as MUEs' behavior and QoS requirements [27].

TABLE 1. SINR parameters.

$\lvert h_{ie}^{n}\rvert$	channel gain between PMU i and SCBS e
p_i	transmission power of PMU i
σ^2	variance of the complex Gaussian thermal noise at the receiver
z_i	channel access action of PMU i

A. SIGNAL-TO-INTERFERENCE-PLUS-NOISE RATIO AND DATA RATE OF THE PMUS

The total interference plus noise measured by each PMU includes interference from MUE-MBS and SUE-SCBS over the same sub-channel, and the additive white Gaussian noise (AWGN). Let γ_i denote the received signal-to-interference-plus-noise ratio (SINR) of PMU i at sub-channel n, which can be calculated as [28]

$$\gamma_i(p_i, z_i) = \frac{|h_{ie}^{n}(z_i)|p_i}{\sigma^2 + \sum_{u\in\mathcal{U}} |h_{ue}^{n}(z_u)|p_u + \sum_{j\in\mathcal{J}} |h_{je}^{n}(z_j)|p_j}. \tag{1}$$

All symbols are explained in Table 1.

The channel gain over sub-channel n can be calculated as $|h_{ie}^{n}| = C\xi_{is}(L_{ie})^{-\alpha}$ [28]. C, ξ_{is}, L_{ie} and α denote the path loss constant, the slow fading component with Nakagami-m distribution, the distance between PMU $i \in \mathcal{M}$ and SCBS e, and the path loss exponent respectively. Nakagami-m distribution is adopted as it applies to a large class of fading channels.

In order to satisfy the QoS for each PMU, different minimal SINR requirements, $\gamma_i^{\min}$, is applied to data transmissions, which is determined according to the state of DER i at time slot t, $g_{i,t}$, expressed as

$$\gamma_i^{\min} = \begin{cases} \gamma_1, & \text{if } g_{i,t} = 0, \\ \gamma_2, & \text{otherwise.} \end{cases} \tag{2}$$

Let $r_{i,k}$ denote the data rate of PMU i at timeslot k which can be calculated as [29]

$$r_{i,k}(p_i, z_i) = \log_2\Big(1 + \gamma_i(p_i, z_i)\Big). \tag{3}$$

B. QUEUE DYNAMICS OF PMUS

Each PMU generates data at each timeslot, and data are divided into packets with the same size. The amount of the packets generated by PMU i at timeslot k is denoted as $B_{i,k}$, with the rate of λ. The generated data is stored in the queue first, and will be transmitted at the next time slot with first in first out (FIFO) behavior. Assume that the size of the buffer is large, so that no data is dropped due to the buffer overflow. The queue length of PMU i at time slot $k+1$ can be defined as

$$Q_{i,k+1} = \max\{0, Q_{i,k} - r_{i,k}(p_i, z_i)\} + B_{i,k}, \tag{4}$$

where $Q_{i,k}$ is denoted as the queue length of PMU i at timeslot k.

C. DELAY AND ENERGY EFFICIENCY MODEL

The average delay of PMU i, $\bar{D}_i$, can be calculated based on the Little's law [30]

$$\bar{D}_i = \frac{\bar{Q}_i}{\bar{T}_i}, \tag{5}$$

where $\bar{Q}_i$ is the average queue length, and $\bar{T}_i$ is the average throughput, and $T_i = \min\{Q_i, r_i(p_i, z_i)\}$.

The total power consumed by PMU i at iteration t, denoted as $P_{i,t}^{\text{Total}}$, can be calculated as

$$P_{i,t}^{\text{Total}} = P_c + p_{i,t}, \tag{6}$$

where P_c is denoted as the circuit power due to signaling and active circuit blocks, and the transmission power $p_{i,t}$. The circuit power can be modeled as the total of a static term and a dynamic term [31], $P_c = VI_{leak} + A_sCfV^2$, where V, I_{leak}, A_s, C and f denote the transistors supply voltage, the leakage current, the fraction of gate actively switching, the circuit capacitance, and the clock frequency respectively. The frequency is assumed to be dynamically scaled with the sum rate, therefore the circuit power can be modeled as [32]

$$P_c = P_s + \beta r_{i,t}, \tag{7}$$

where P_s denotes the static term and β is a constant representing dynamic power consumption per unit data rate. In this work, the circuit power is calculated when PMUs generate data until the data arrive at the control center.

Energy efficiency is usually defined as information bit per unit of energy, which corresponds to the ratio of the data rate to the unit power consumption, which can be calculated as [33],

$$EE_k = \frac{\sum_{i=1}^{M} r_{i,k}(p_i, z_i)}{\sum_{i=1}^{M} P_i^{\text{Total}}}. \tag{8}$$

IV. A PROPOSED DEEP REINFORCEMENT LEARNING APPROACH FOR THE CHANNEL ACCESS AND POWER CONTROL SCHEME

The goal of the DRL approach is to ensure that no PMU receives SINR falls below the threshold $\gamma_i^{\min}$, for successful transmissions, $\gamma_i \geq \gamma_i^{\min}, \forall i \in \mathcal{I}$ and the interference caused by PMUs, $h_{ie}p_i(z_i)$, is not greater than an interference threshold I_u^{th}, $h_{ie}p_i(z_i) \leq I_u^{\text{th}}, \forall u \in \mathcal{U}$, to protect the QoS of MUEs.

Almost all RL problems can be formulated as Markov decision process (MDP) as an MDP can describe the environment for RL, which is fully observable. Therefore, to adopt a DRL approach, first, the elements in MDP need to be defined. The goal of the MDP in a RL problem is to maximize the earned rewards [20], [34].

A. MARKOV DECISION PROCESS ELEMENTS

Let $\mathcal{S}$ and $\mathcal{A}$ denote the set of the states and the actions for the agent, respectively. PMU i senses the state $s_{i,t} \in \mathcal{S}$ and selects an action $a_{i,t} \in \mathcal{A}$ at each timeslot t. Based on the action taken, the environment makes a transition to a new state,

$s_{i,t+1} \in \mathcal{S}$ according to probability $Pr(s_{i,t+1}|s_{i,t}, a_{i,t})$ and generates a reward, $R_{i,t}(s_{i,t}, a_{i,t})$ to the agent. In this paper, a DRL approach is proposed to obtain optimal policy for channel access and power control in HetNets. However, in order to utilize the DRL technique for the PMUs, the state space, the action space and the reward function need to be defined.

1) STATE SPACE

The environment system state is defined based on local observations of the PMUs, therefore at timeslot t, the state $s_{i,t}$ observed by PMU $i \in \mathcal{M}$ can be expressed as follows.

$$s_{i,t} = I_{i,t}, \zeta_{i,t}, \tag{9}$$

where $I_{i,t} \in \{0, 1\}$ indicates whether the received SINR of PMU i, $\gamma_{i,t}$, is above or below the minimum SINR, $\gamma_i^{\min}$, which is expressed as follows.

$$I_{i,t} = \begin{cases} 1, & \text{if } \gamma_{i,t}(p_{i,t}, z_{i,t}) \geq \gamma_i^{\min}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

On the other hand, $\zeta_{i,t}$ denotes whether the interference caused by PMU i over sub-channel n occupied by MUE u is above or below the interference threshold, such that

$$\zeta_{i,t} = \begin{cases} 1, & \text{if } h_{ie,t}^{n}(z_{i,t})p_{i,t} \leq I_u^{\text{th}}, \quad \forall u \in \mathcal{U} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The state space of the whole system at timeslot t is expressed as $\mathcal{S}_t = \{s_{i,t}, \cdots, s_{M,t}\}$.

2) ACTION SPACE

An action performed by each PMU at each timeslot considers discrete changes in the channel access, as well as the transmission power level, therefore the action set of PMU i is denoted as $\mathcal{A}_i = [\mathcal{Z}_i, \mathcal{P}_i]$, where $\mathcal{Z}_i = [Z_1, Z_2, \cdots, Z_N]$ and $\mathcal{P}_i = [P_1, P_2, \cdots, P^{\max}]$. The action set defines a discrete set of available actions that the PMU can perform at each timeslot. The action is selected to maximize the reward, by considering the minimum SINR requirement and interference to the MUE. The PMU first determines the $\gamma_i^{\min}$, then selects a set of sub-channel and transmission power that satisfies its delay as well as maximizes energy efficiency.

3) REWARD FUNCTION

When a distributed scheme is implemented in HetNets, one of the concern is the reward. A higher SINR at the PMU will result in lower delay, however achieving a high SINR requires the PMU to transmit at a high power level, causing more power consumption as well as increasing the magnitude of interference to other MUEs. Therefore, the energy efficiency of the PMUs is selected as the reward function, expressed as [33]

$$R_{i,t}(p_{i,t}, z_{i,t}) = r_{i,t}(p_{i,t}, z_{i,t})/P_{i,t}^{\text{Total}}, \tag{12}$$

The reward $R_{i,t}(s_{i,t}, a_{i,t})$ of PMU i in state $s_{i,t}$ is the immediate return when action $a_{i,t}$ is executed, which is formulated as [27]

$$R_{i,t}(s_{i,t}, a_{i,t}) = \begin{cases} R_{i,t}(p_{i,t}, z_{i,t}), & \text{if } I_{i,t} = 1 \text{ and } \zeta_{i,t} = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

In particular, the reward is a return of selecting channel $z_{i,t}(a_{i,t})$ and power level $p_{i,t}(a_{i,t})$ in state $s_{i,t}$ that ensures the transmission delay constraints and/or achieves energy efficiency.

B. Q-LEARNING FOR PMU

The goal of RL approach is to improve the PMU's decision-making policy, π over time. The policy π, can be defined as a mapping from environment states to probability distribution over actions. However, learning a policy is difficult, hence, some RL approaches attempt to learn the policy indirectly [34]. This can be done by learning the optimal value function (either a state-value function or an action-value function). Depending on the function chosen, the agent will learn the value of being in a specific state (state-value function) or being in a specific state and taking certain action (action-value function). Therefore, by learning the optimal value function, the optimal policy, π^*, can be inferred [34].

The task of the PMUs is to learn the optimal policy, π^* that maximizes the total expected discounted reward over infinite steps, expressed as

$$V^{\pi}(s_{i,t}) = \sum_{t=1}^{T} \phi^{t-1} R_{i,t}, \tag{14}$$

in which ϕ and T are the discounted factor and the time where the goal state, where the action remains unchanged is obtained respectively. Therefore, the task becomes learning an optimal policy π^* that can maximize V^{π}, which can be described as follows [18]

$$\pi^* = \arg\max_{\pi} V^{\pi}(s_t). \tag{15}$$

It is difficult to learn π^* in (15), therefore Q-learning approach is adopted to solve the equation. In Q-learning, an action-value function, also known as Q function, is introduced to evaluate the expected discounted cumulative reward after execute action $a_{i,t}$ in state $s_{i,t}$. The optimal policy can be constructed by selecting the highest value in each state when an action function is learned. In Q-learning, the PMU tries to update the Q function using the update rule known as Bellman equation [27]

$$\begin{aligned} Q(s_{i,t}, a_{i,t}) = {} & Q(s_{i,t}, a_{i,t}) + \alpha R_{i,t}(s_{i,t}, a_{i,t}) \\ & + \phi \max_{a_{i,t+1}} Q(s_{i,t+1}, a_{i,t+1}) - Q(s_{i,t}, a_{i,t}), \end{aligned} \tag{16}$$

where α is the learning rate.

Equation (16) has been proven to converge to the optimal action-value function, which is defined as the maximum expected discounted cumulative reward by following any policy, after executing action $a_{i,t}$ in state $s_{i,t}$ [18]. In Q-learning, the number of states is finite and the action-value function

is estimated separately for each state, forming a Q-table in which the rows represent the states and the columns represent the possible actions. When the Q-table converges, the PMU can select an action with the highest $Q(s_{i,t}, a_{i,t})$ value as the optimal action in state $s_{i,t}$.

However, due to the curse of dimensionality in HetNets, the Q-learning method is impractical for the problem, as it needs to store a value for every possible state-action pair in the Q-Table, requiring a lot of memory and time to converge [34]. In order to overcome this issue, a technique, known as value function approximation, is introduced, in which the Q-Table is now represented and its values are estimated by a function. This function is learned online by the agent's interaction with the environment and it can be of any kind, such as linear or logistic regression, neural networks, or deep neural networks [34]. Based on this technique, a deep Q-learning (DQN) approach is proposed in which a DNN is utilized to approximate the action-value function, now represented as $Q(s_{i,t}, a_{i,t}; \theta)$, where θ represents the weights learned by the DNN.

C. DEEP REINFORCEMENT LEARNING ALGORITHM FOR CHANNEL ACCESS AND POWER CONTROL SCHEME

When Q-Learning is combined with a DNN, DQN is created. DQN, which is another term of DRL utilizes a DNN to derive the correlation between state-action pairs $(s_{i,t}, a_{i,t})$ then estimates its value function $Q(s_{i,t}, a_{i,t}; \theta_{i,t})$ [16]. However, when combining a DNN with Q-Learning, several problems regarding convergence and stability arise [19]. As such in [16], the authors proposed two mechanisms to overcome these issues. First, a technique known as experience replay was added, in which the agent's experiences with the environment are stored in a memory and utilized, via a random mini-batch process to train the neural network. The second modification is to use two separate neural networks, one which is constantly evaluated and updated according to the agent's experience, and another one, a target network, in which the weights are periodically updated. In addition to this, an online training mechanism is devised, so that based on the agent's interaction with the environment and its observations, the values of the action-value function can be learned. The training data used to train the Q-network for each PMU is generated as follows.

Given $s_{i,t}$ at iteration t for PMU i, an action $a_{i,t}$ is randomly selected with probability ε_t, or selected with the largest output $Q(s_{i,t}, a_{i,t}; \theta_0)$ (following the ϵ-greedy policy), where θ_0 denotes the weights of the DNN at the current iteration. After taking an action $a_{i,t}$, PMU i receives a reward $R_{i,t}$ and observes a new state $s_{i,t+1}$. This transition $d_{i,t} \triangleq \{s_{i,t}, a_{i,t}, R_{i,t}, s_{i,t+1}\}$, is stored in the replay memory D. The training of the Q-network begins when D has collected a sufficient number of transitions, assume $O = 300$ transitions. Specifically, a minibatch of transitions $\{d_w | w \in \Omega_t\}$ from D is randomly selected, and the Q-network can be trained by adjusting the parameter θ to minimize the loss function,

Algorithm 1 DRL Training for Channel Access and Power Control Scheme

1: Input: replay memory D with buffer capacity O, training steps T, target network learning rate α.
2: Initialize $Q(s, a; \theta_0)$ with random weights θ_0
3: Initialize $a_{i,1}$, then obtain $s_{i,1}$
4: **for all** $t = 1, T$ **do**
5: With probability ε_t, select a random action $a_{i,t}$ otherwise $a_{i,t} = \arg\max_a Q(s_{i,t}, a; \theta_0)$.
6: Execute action $a_{i,t}$ and observe reward $R_{i,t}$ and obtain $s_{i,t+1}$
7: Store transition $d_{i,t} \triangleq \{s_{i,t}, a_{i,t}, R_{i,t}, s_{i,t+1}\}$ in D.
8: **if** $t \geq O$ **then**
9: Sample a random minibatch of transitions $\{d_w | w \in \Omega_t\}$ from D, where the indexes of Ω_t are uniformly selected randomly
10: Update θ by minimizing the loss function (17), in which targets Q'_w are given by (18)
11: Set $\theta_0 = \arg\min_\theta L(\theta)$
12: **end if**
13: **end for**
14: Output: $Q(s, a, \theta)$

expressed as follows

$$L(\theta) \triangleq \frac{1}{\Omega_t} \sum_{w \in \Omega_t} \left(Q'_{i,w} - Q(s_{i,w}, a_{i,w}; \theta)\right)^2, \qquad (17)$$

in which Ω_t denotes the index set of the random minibatch used at the t-th iteration, and $Q'_{i,w}$ is a value estimated using a Bellman equation, by fixing set of weights from the previous iterations of the learning procedure.

The target of DRL can be expressed as follows

$$Q'_{i,w} = R_{i,w} + \phi \max_{a'} Q(s_{i,w+1}, a'; \theta_0), \quad \forall w \in \Omega_t, \qquad (18)$$

where θ_0 is the set of fixed weights from previous DNN iterations. In DRL, the targets are updated as the weight θ is refined, which is different from traditional supervised learning.

The algorithm of DRL training for channel access and power control is described in Algorithm 1. In the training process, a PMU achieves a goal state at s_t if the action remains unchanged at the next state s_{t+1}. Therefore, it is not difficult to prove that the next state s_{t+1} is also a goal state. Assume that once s_t achieves a goal state, it stays at the goal state until the transmission is done. Then, the policy has been converged at this rate, and the largest estimated value $Q(s, a, \theta^*)$ is obtained. After the training process, for each state, the PMU selects an action which yields the largest estimated value $Q(s, a, \theta^*), p_{i,t}, z_{i,t} = \max_a Q(s, a, \theta^*)$.

V. SIMULATION RESULTS AND DISCUSSION

The performance of the proposed scheme is evaluated using Tensorflow, and the same environment in [28] is considered. System parameters are explained and experimental results are discussed in this section.

TABLE 2. Simulation parameters.

Macrocell radius	400 m
Small cell radius	50 m
Number of MUEs	30
Number of SUEs	15
Number of PMU	3
Number of sub-channels	30
Data rate (λ)	60 packets/s
Minimum SINR requirement (γ_1, γ_2)	15, 35 dB
Interference threshold of MUEs (I_u^{th})	10^-6
Noise power density (σ^2)	-174 dBm/Hz
Path loss constant (C^s)	10^{-2}
Path loss component (α)	4.8
Bandwidth of each sub-channel	180 kHz
Transmission power of MUEs and SUEs	19 dBm
Maximum transmission power of PMU	19 dBm
PMU circuit power	20 dBm
Number of hidden layers	3
Activation function	ReLUs
Size of minibatch	256
Number of iterations	$35x10^3$
Weight update	Adam algorithm
Replay memory	400 transitions
Buffer capacity	300 transitions

A. SIMULATION PARAMETERS

There are 3 PMUs, 13 SUEs and 30 MUEs uniformly distributed in a cell with 400 m radius, located in a rural area. Each PMU generates a typical packet size of 52 Bytes with the rate of $\lambda = 60$ packets/s [35]. The length of each timeslot is 1 ms. In the simulation, each PMU selects a sub-channel from a predefined set $\mathcal{Z} = \{1, 2, \cdots, 30\}$ and the transmission power (in dBm) is selected from set $\mathcal{P} = \{14, 15, \cdots, 19\}$. Regarding the DRL parameters, each PMU is trained in the DNN to approximate its action-value function. The DRL consists of three hidden layers with 256, 256 and 512 neurons on each layer respectively. The first two hidden layers use rectified linear units (ReLUs) as the activation functions, while the last layer uses a tanh function. The weights θ are updated by adopting a recently proposed adaptive moment estimation (Adam) algorithm [36]. The reason for this is because it requires only first-order gradients with small memory requirement to achieve the optimum [36]. The PMUs explore new actions with the probability from 0.8 to 0.05 between iterations, in which at iteration t, the probability can be expressed as $\varepsilon_t = 0.8(1 - t/T)$. A detailed list of simulation parameters is given in Table 2.

In this paper, three different decision-making policies are used for comparison, which are explained as follows

- DRL policy: the action is selected based on Algorithm 1.
- Myopic policy: this policy selects the action with maximum expected immediate reward and ignores the impact of the current action on the future reward [37].
- Gittin policy: this policy calculates the Gittin index for each action, which is the accumulated reward per unit time and selects the action with the maximum value [38].

The Myopic policy and the Gittin policy are easy to implement but both policies require prior knowledge of the system dynamics, which is not easy to obtain beforehand [21].

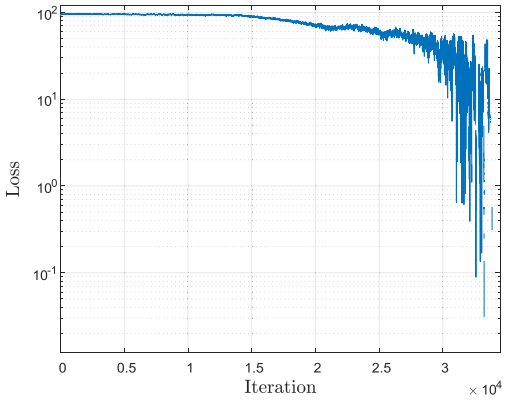


FIGURE 2. Loss of Q function with various iterations during the training process.

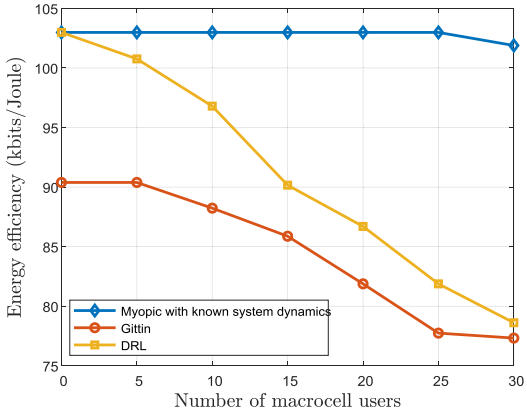


FIGURE 3. Energy efficiency comparison among three policies for various number of users.

B. PERFORMANCE OF DRL ALGORITHM

We conduct a simulation to evaluate performance of the proposed algorithm for $35k$ independent runs during the training process. The performance of the DRL algorithm is evaluated in terms of loss Q function which is calculated as in (17). In general, Fig. 2 shows that the loss of the Q function decreases as the number of iterations increases and becomes constant at the lowest loss function value after $34k$ training iterations. This shows that the proposed algorithm can successfully converge and the PMU can make the optimal decision given any system state.

C. THE IMPACT OF NUMBER OF USERS

The impact of number of macrocell users on the performance of all three policies when the minimum SINR requirement of PMU is 15 dB is investigated. Fig. 3, Fig. 4 and Fig. 5 compare the energy efficiency, average delay and power consumed by all three polices respectively. The results show that the Myopic policy with known system dynamics achieves the best energy efficiency but the worse average delay for all number of users. The reason for that is because the aim of the Myopic policy is to maximize the immediate reward, which is the energy efficiency, therefore the policy consumed

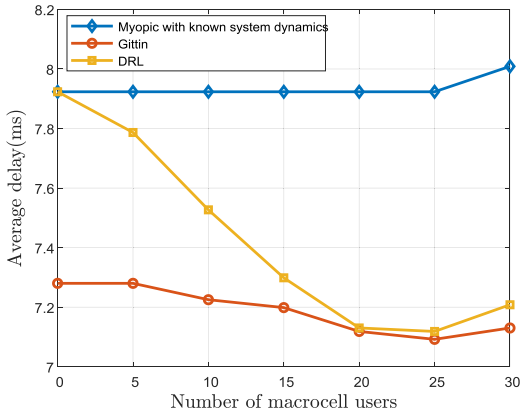


FIGURE 4. Average delay comparison among three policies for various number of users.

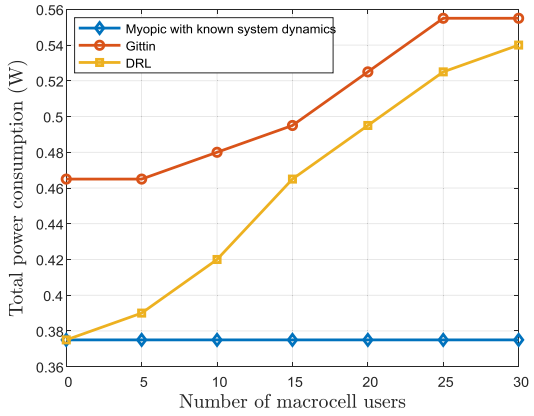


FIGURE 5. Power consumption comparison among three policies for various number of users.

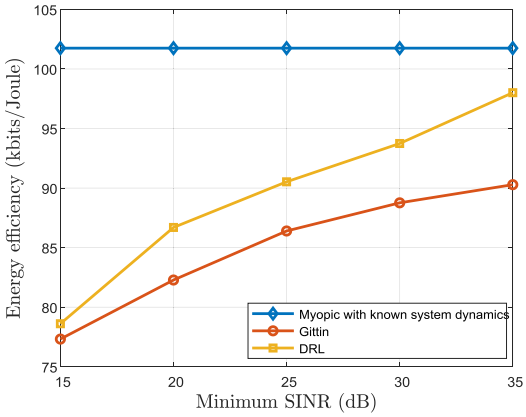


FIGURE 6. Energy efficiency comparison among two policies with varying minimum SINRs.

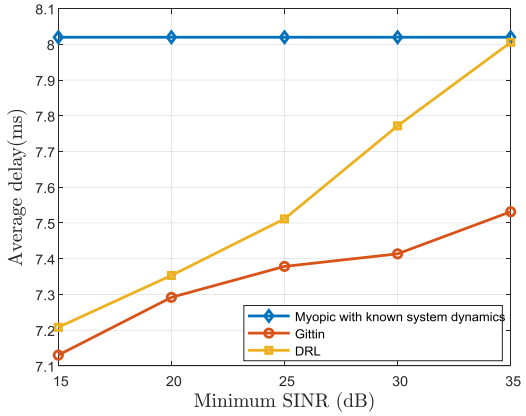


FIGURE 7. Average delay comparison among two policies with varying minimum SINRs.

the lowest power, which result in lowest data rate yielding low delay and high energy efficiency. Moreover, this policy has a constant energy efficiency and average delay between 0 to 25 users and becomes worse after 30 users. The reason for that is because there are empty sub-channels that are not shared with MUEs at 0 to 25 users, therefore the policy selects the empty sub-channel, while at 30 users, all sub-channel are occupied by MUEs and must be shared with PMUs which increase the interference, therefore the performance decreases.

On the other hand, the energy efficiency of the DRL policy and the Gittin policy decreases as the number of users increases because there is less chance to find a good action when more sub-channels are shared with MUEs, consequently decreasing the energy efficiency. The average delay of both DRL and Gittin policies decrease from 0 to 25 users since energy efficiency is decreasing due to high power consumption, increasing the data rate. However, the delay increases at 30 user because at this number, all sub-channel are occupied, yielding the highest interference, which increases the delay when maximizing the energy efficiency. Additionally, the results show that the DRL policy can learn the system dynamics and achieve good performance as well as outperforms the Gittin policy even without knowledge of the system dynamics beforehand.

D. THE IMPACT OF MINIMUM SINR

We conduct simulations to investigate the impact of minimum SINR requirements on the performance of all three policies when all sub-channels are shared with MUEs. Fig. 6 shows that the energy efficiency of DRL and Gittin policies are getting better as the minimum SINR requirement increases. The reason for that is because as the constraints get more stringent, there is more chance to select a good action since the actions that failed to meet the constraints have been eliminated. On the other hand, the Myopic policy with knowledge of the system dynamics has the highest and constant energy efficiency for all constraints because with this knowledge, the policy is able to obtain the best action in the very beginning. However, the average delay of this policy, as shown in Fig. 7 is the highest and at 35 dB, and this policy fails to meet the delay constraint, which is 8 ms. The result also shows that average delay of the DRL and Gittin policies increase as the minimum SINR requirement increases since the energy efficiency is maximized as the minimum SINR

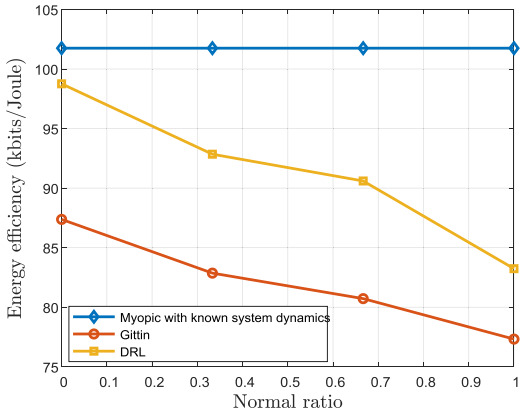


FIGURE 8. Energy efficiency comparison among three schemes with varying normal ratio.

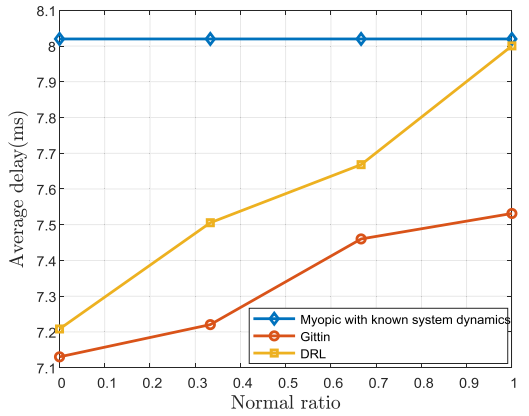


FIGURE 9. Average delay comparison among three schemes with varying normal ratios.

requirement increases, hence the data rate decreases, resulting in higher delay. However, both policies are still able to meet the delay constraints in all minimum SINR requirements. Moreover, the results show that the DRL policy outperforms the Gittin policy at all minimum SINR requirements because in the DRL policy, as the constraints become more stringent, there is more chance to select a good action, therefore the learning process becomes easier, so the PMU is able to find the optimal policy easily and quickly.

E. THE IMPACT OF NORMAL RATIO

We study the impact of normal ratio on the performance of all three policies. Normal ratio is defined as the ratio of number of PMUs observing DERs in normal states to the total number of PMUs in the cell. In this work, only 3 PMUs are located in the cell, yielding the gap of 1/3 between normal ratios. Fig. 8 and Fig. 9 show that energy efficiency and average delay of the DRL policy and the Gittin policy become worse as the normal ratio increases. The reason for that is lesser PMUs which are in abnormal or restorative states, the constraints get more lenient which is the minimum SINR requirements are lower, therefore there is less chance to find a good action, hence decreasing the energy efficiency. However, both policies are still able to meet the delay requirement as the normal ratio increases. On the other hand, the performance of the Myopic policy is constant even when the normal ratio increases due to the fact that the different minimum SINR requirements of PMUs do not affect the performance of the Myopic policy. Moreover, the results show that the DRL policy outperforms the Gittin policy in all normal ratios. This shows that the DRL policy can be applied in more complex situations where more PMUs are involved in the cell.

VI. CONCLUSION

This paper studied HetNets for simultaneous transmissions of smart grid NANs data, in particular the PMU data. An intelligent channel access and power control scheme was proposed to maximize energy efficiency in HetNets as well as satisfy the delay constraints. A DRL approach was exploited to obtain optimal policy that maximizes the discounted reward and enable successful data transmission with the considerations on the minimum SINR requirements of PMUs and also the interference caused by PMUs to the MUEs and other SUEs. In DRL, each PMU was trained using DQN-based intelligent channel access and power control algorithm, where the data of the environment were extracted and predictive guesses about new data were made using a cascade of layers of processing units. After the training, the PMU selects an action that maximizes the reward function. Simulation results showed that the PMUs able to learn the system dynamics and obtain optimal policy in any given state. Additionally, the DRL policy provides excellent performance in different number of users, minimum SINR requirements and normal ratios compared to the Gittin policy even without knowledge of the system dynamics beforehand. One interesting topic which can be done in future is the interdependencies between communication and power networks, specifically study on the impact of PMU's end-to-end delay in HetNets to the total power loss of the grid.

REFERENCES

[1] H. Farhangi, "The path of the smart grid," *IEEE Power Energy Mag.*, vol. 8, no. 1, pp. 18–28, Jan./Feb. 2010.

[2] T. Sauter and M. Lobashov, "End-to-end communication architecture for smart grids," *IEEE Trans. Ind. Electron.*, vol. 58, no. 4, pp. 1218–1228, Apr. 2011.

[3] F. H. Fesharaki, R.-A. Hooshmand, and A. Khodabakhshian, "Simultaneous optimal design of measurement and communication infrastructures in hierarchical structured WAMS," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 312–319, Jan. 2014.

[4] N. Kayastha, D. Niyato, E. Hossain, and Z. Han, "Smart grid sensor data collection, communication, and networking: A tutorial," *Wireless Commun. Mobile Comput.*, vol. 14, no. 11, pp. 1055–1087, 2014.

[5] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Trans. Ind. Electron.*, vol. 57, no. 10, pp. 3557–3564, Oct. 2010.

[6] M. Qiu, W. Gao, M. Chen, J.-W. Niu, and L. Zhang, "Energy efficient security algorithm for power grid wide area monitoring system," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 715–723, Dec. 2011.

[7] Y. Cao, T. Jiang, M. He, and J. Zhang, "Device-to-device communications for energy management: A smart grid case," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 190–201, Jan. 2016.

[8] N. Xia, H.-H. Chen, and C.-S. Yang, "Radio resource management in machine-to-machine communications—A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 791–828, 1st Quart., 2018.

[9] A. Suzdalenko and I. Galkin, "Case study on using non-intrusive load monitoring system with renewable energy sources in intelligent grid applications," in *Proc. Int. Conf.-Workshop Compat. Power Electron.*, Jun. 2013, pp. 115–119.

[10] X. Lu, W. Wang, and J. Ma, "An empirical study of communication infrastructures towards the smart grid: Design, implementation, and evaluation," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 170–183, Mar. 2013.

[11] M. Moltafet, P. Azmi, N. Mokari, M. R. Javan, and A. Mokdad, "Optimal and fair energy efficient resource allocation for energy harvesting-enabled-PD-NOMA-based HetNets," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2054–2067, Mar. 2018.

[12] K. Hammad, A. Moubayed, S. L. Primak, and A. Shami, "QoS-aware energy and jitter-efficient downlink predictive scheduler for heterogeneous traffic LTE networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 6, pp. 1411–1428, Jun. 2018.

[13] L. Tang, W. Wang, Y. Wang, and Q. Chen, "An energy-saving algorithm with joint user association, clustering, and on/off strategies in dense heterogeneous networks," *IEEE Access*, vol. 5, pp. 12988–13000, 2017.

[14] C. Shen, C. Tekin, and M. van der Schaar, "A non-stochastic learning approach to energy efficient mobility management," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3854–3868, Dec. 2016.

[15] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digit. Commun. Netw.*, vol. 1, no. 1, pp. 1–19, 2015.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[17] B. Zhang, C. H. Liu, J. Tang, Z. Xu, J. Ma, and W. Wang, "Learning-based energy-efficient data collection by unmanned vehicles in smart cities," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1666–1676, Apr. 2018.

[18] X. Li, J. Fang, W. Cheng, H. Duan, Z. Chen, and H. Li, "Intelligent power control for spectrum sharing in cognitive radios: A deep reinforcement learning approach," *IEEE Access*, vol. 6, pp. 25463–25473, 2018.

[19] B. Cao, L. Zhang, Y. Li, D. Feng, and W. Cao, "Intelligent offloading in multi-access edge computing: A state-of-the-art review and framework," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 56–62, Mar. 2019.

[20] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J. Oh, "Semisupervised deep reinforcement learning in support of IoT and smart city services," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 624–635, Apr. 2018.

[21] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 257–265, Jun. 2018.

[22] B. Cao, S. Xia, J. Han, and Y. Li, "A distributed game methodology for crowdsensing in uncertain wireless scenario," *IEEE Trans. Mobile Comput.*, to be published.

[23] H. Gharavi and B. Hu, "Synchrophasor sensor networks for grid communication and protection," *Proc. IEEE*, vol. 105, no. 7, pp. 1408–1428, Jul. 2017.

[24] T. E. D. Liacco, "The adaptive reliability control system," *IEEE Trans. Power App. Syst.*, vol. PAS-86, no. 5, pp. 517–531, May 1967.

[25] K. V. Katsaros, B. Yang, W. K. Chai, and G. Pavlou, "Low latency communication infrastructure for synchrophasor applications in distribution networks," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2014, pp. 392–397.

[26] P. Popovski *et al.*, "Scenarios requirements and KPIs for 5G mobile and wireless system," METIS Project, Mobile Wireless Commun. Enablers Twenty-Twenty Inf. Soc., Tech. Rep. ICT-317669-METIS D, 2013, vol. 1.

[27] X. Chen, Z. Zhao, and H. Zhang, "Stochastic power adaptation with multiagent reinforcement learning for cognitive wireless mesh networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2155–2166, Nov. 2013.

[28] H. Dai, Y. Huang, R. Zhao, J. Wang, and L. Yang, "Resource optimization for device-to-device and small cell uplink communications underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1187–1201, Feb. 2018.

[29] S. Samarakoon, M. Bennis, W. Saad, and M. Latva-Aho, "Backhaul-aware interference management in the uplink of wireless small cell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5813–5825, Nov. 2013.

[30] L. Lei, Y. Kuang, N. Cheng, X. S. Shen, Z. Zhong, and C. Lin, "Delay-optimal dynamic mode selection and resource allocation in device-to-device communications—Part I: Optimal policy," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3474–3490, May 2016.

[31] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3767–3778, Dec. 2012.

[32] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," in *Proc. Global Telecommun. Conf. GLOBECOM*, Dec. 2010, pp. 1–6.

[33] B. Yang, K. V. Katsaros, W. K. Chai, and G. Pavlou, "Cost-efficient low latency communication infrastructure for synchrophasor applications in smart grids," *IEEE Syst. J.*, vol. 12, no. 1, pp. 948–958, Mar. 2018.

[34] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

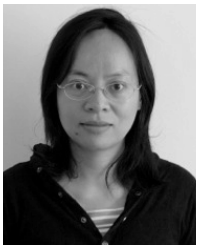
[35] M. Kuzlu, M. Pipattanasomporn, and M. Rahman, "Communication network requirements for major smart grid applications in HAN, NAN and WAN," *Comput. Netw.*, vol. 67, pp. 74–88, Jul. 2014.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[37] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5547–5567, Nov. 2010.

[38] J. Ai and A. A. Abouzeid, "Opportunistic spectrum access based on a constrained multi-armed bandit formulation," *J. Commun. Netw.*, vol. 11, no. 2, pp. 134–147, Apr. 2009.

FAUZUN ABDULLAH ASUHAIMI received the bachelor's degree (Hons.) in communication engineering from International Islamic University Malaysia, in 2012, and the master's degree in electrical engineering (telecommunications) from the University Technology of Malaysia, in 2014. She is currently pursuing the Ph.D. degree with the University of Glasgow, Glasgow, U.K. Her research interests include the areas of cellular technology, 5G communications, and smart grids.



SHENGRONG BU received the Ph.D. degree in electrical and computer engineering from Carleton University, in 2012.

She held a research position with Huawei Technologies Canada Inc., Ottawa, as a NSERC IRDF, until 2014. She is currently a Lecturer (Assistant Professor equivalent) with the School of Engineering, University of Glasgow, Scotland. Her research interests include energy efficient networks, smart grids, big data analytics, wireless networks, wireless network security, cloud computing, game theory, and stochastic optimization. She received the Best Paper Awards at the International IEEE Conference on Industrial Informatics (INDIN 2005), the IEEE Global Communication Conference (Globecom 2012), and received one of the Best 50 Papers Award at the IEEE GLOBECOM'2014. She was also awarded the NSERC PDF Fellowship (Rank: 1st in Electrical Engineering, Canada), in 2014. She has served as an Associate Editor for the *Springer Wireless Networks* and also an Editor for the IEEE TCGCC NewsLetters. She was a TPC Co-Chair for six international workshops or conference symposiums, and served duties as the N2Women Mentoring Co-Chair. She has sat on the TPC for more than 20 leading international conferences and workshops and served as a Reviewer for more than ten leading journals.

PAULO VALENTE KLAINE received the B.Eng. degree in electrical and electronic engineering from the Federal University of Technology–Paraná (UTFPR), Brazil, in 2014, and the M.Sc. degree in mobile communications systems from the University of Surrey, Guildford, U.K., in 2015, both with distinction. He is currently pursuing the Ph.D. degree with the School of Engineering, University of Glasgow. In 2016, he spent the first year of his Ph.D., working with the 5G Innovation Centre (5GIC), University of Surrey. His main interests include self-organizing cellular networks and the application of machine learning algorithms in wireless networks.

MUHAMMAD ALI IMRAN received the M.Sc. (Hons.) and Ph.D. degrees from Imperial College London, U.K., in 2002 and 2007, respectively. He has over 18 years of combined academic and industry experience, working primarily in the research area of cellular communication systems. He is currently an Affiliate Professor with The University of Oklahoma, Norman, OK, USA, and a Visiting Professor with the 5G Innovation Centre, University of Surrey, U.K. He is also the Vice Dean of the Glasgow College, UESTC, and a Professor of communication systems with the School of Engineering, University of Glasgow. He has led a number of multimillion-funded international research projects encompassing the areas of energy efficiency, fundamental performance limits, sensor networks, and self-organizing cellular networks. He also led the new physical layer work area for 5G Innovation Centre at Surrey. He has a global collaborative research network spanning both academia and key industrial players in the field of wireless communications. He has supervised more than 30 successful Ph.D. graduates. He has been awarded for his excellence in academic achievements, conferred by the President of Pakistan. He received the IEEE Comsoc's Fred Ellersick Award, in 2014, the FEPS Learning and Teaching Award, in 2014, and the Sentinel of Science Award, in 2016. He was twice nominated for the Tony Jean's Inspirational Teaching Award. He was also a shortlisted finalist for the Wharton-QS Stars Awards, in 2014, the QS Stars Reimagine Education Award for innovative teaching and VC's learning, in 2016, and the Teaching Award in the University of Surrey. He has given an invited TEDx talk (2015) and more than ten plenary talks, several tutorials and seminars in international conferences, events, and other institutions. He has taught on international short courses in USA and China. He is the Co-Founder of the IEEE Workshop BackNets 2015 and chaired several tracks/workshops of international conferences. He is an Associate Editor of the IEEE Communications Letters, IEEE Open Access, and the *IET Communications Journal*, and has served as a Guest Editor for many prestigious international journals.